

Wherefore, what is claimed is:

1. A computer-implemented process for using feature selection to obtain a strong classifier from a combination of weak classifiers, comprising using a computer to perform the following process actions:

(a) inputting a set of training examples, a prescribed maximum number of weak classifiers, a cost function capable of measuring the overall cost, and an acceptable maximum cost;

(b) computing a set of weak classifiers, each classifier being associated to a particular feature of the training examples,

(c) determining which of the set of weak classifiers is the most significant classifier;

(d) adding said most significant classifier to a current set of optimal weak classifiers;

(e) determining which of the current set of optimal weak classifiers is the least significant classifier;

(f) computing the overall cost for the current set of optimal weak classifiers using the cost function;

(g) conditionally removing the least significant classifier for the current set of optimal weak classifiers;

(h) computing the overall cost for the current set of optimal weak classifiers less the least significant classifier using the cost function;

(i) determining whether the removal of the least significant classifier results in a lower overall cost;

(j) whenever it is determined that the removal of the least significant classifier results in a lower overall cost, eliminating the least significant classifier;

(k) recomputing each classifier in the current set of optimal weak classifiers associated with a feature added subsequent to the eliminated classifier while keeping the earlier optimal weak classifiers unchanged;

(l) repeat actions (f) through (k) until it is determined the removal of the least significant classifier does not result in a lower overall cost and then reinstating the last identified least significant classifier to the current set of optimal weak classifiers;

(m) determining if the number of weak classifiers in the current set of optimal weak classifiers equals the prescribed maximum number of weak classifiers or the last computed overall cost for the current set of optimal weak classifiers is less than the acceptable maximum cost; and

(n) whenever it is determined that the number of weak classifiers in the current set of optimal weak classifiers does not equal the prescribed maximum number of weak classifiers and the last computed overall cost for the current set of optimal weak classifiers exceeds the acceptable maximum cost, repeating actions (c) through (m) until it is determined that the number of weak classifiers in the current set of optimal weak classifiers does equal the prescribed maximum number of weak classifiers or the last computed overall cost for the current set of optimal weak classifiers becomes less than the maximum allowable cost, then outputting the sum of the individual weak classifiers as the trained strong classifier.

2. The process of Claim 1 wherein the process action of computing each classifier of a set of weak classifiers comprises the process action of deriving each classifier based on a histogram of a scalar value feature for face training examples and a histogram of a scalar value feature for the non-face training examples.

3. The process of Claim 1 wherein the most significant classifier includes the feature that is the most likely to predict whether a training example matches the classification of a particular classifier.

4. The process of Claim 1 wherein the set of weak classifiers are designed to classify whether a training example is a face or non-face.

5. The process of Claim 1 wherein the set of weak classifiers is designed to classify a training example as a text type.

6. The process of Claim 1 wherein the set of weak classifiers is designed to classify a training example as a type of document.

7. The process of Claim 1 wherein the set of weak classifiers is designed to classify a training example as a speech pattern.

8. The process of Claim 1 wherein the set of weak classifiers is designed to classify a training example as a type of medical condition.

9. The process of Claim 1 wherein a weak classifier  $h_j^*(x)$  is computed as

$$h_j^*(x) = \frac{1}{2} \left[ \log \frac{P_j(x | y = +1, w)}{P_j(x | y = -1, w)} + \log \frac{P(y = +1)}{P(y = -1)} \right]$$

wherein the probability densities of a feature  $j$  for a sub-sample  $x$  of a training example is denoted by  $P_j(x | y = +1)$  for a sought pattern and  $P_j(x | y = -1)$  for a non-sought pattern and the normalized weights are denoted by  $w$ .

10. The process of Claim 9 wherein the probability density for a sought pattern and the probability density for a non-sought pattern can be estimated using the histograms resulting from weighted voting of the training examples.

11. The process of Claim 9 wherein the process action of determining which of the set of weak classifiers is the most significant classifier comprises defining the most

significant classifier  $h_M(x)$  as,  $h_M(x) = \arg \min_{h^* \in H_w^*} \sum e^{-y_i[h(x_i) + h^*(x_i)]}$ , wherein

$H_w^* = \{h_j^*(x) | \forall_j\}$ ,  $h(x) = \sum_{m=1}^{M-1} h_m(x)$ , and  $M$  is the total number of weak classifiers in the set of weak classifiers.

12. The process of Claim 9 wherein the process action of determining which of the set of weak classifiers is the least significant classifier comprises defining the least significant classifier  $h'(x)$  as,  $h' = \arg \min_{h \in H_M} J(H_M - h)$  where  $H_M$  denotes the strong classifier built upon the current set  $H_M$  of selected weak classifiers.

13. The process of Claim 1 wherein the process action of computing the overall cost comprises computing the overall cost  $J(h(x))$  as

$J(h(x)) = \sum_i e^{-y_i h(x_i)}$  wherein  $y=+1$  for a sought pattern and  $y=-1$  for a nonsought pattern and  $h(x_i)$  is a weak classifier in the set of weak classifiers.

14. The process of Claim 1 wherein outputting the sum of the individual weak classifiers as the trained strong classifier comprises outputting the sum  $H(x)$  as  $H(x) = \text{sign}[\sum_{m=1}^M h_m(x)]$  wherein  $M$  is the total number of weak classifiers in the set of weak classifiers  $h_m(x)$  is a weak classifier in the current set of weak classifiers.

15. A system for detecting a person's face in an input image and identifying a face pose range into which the face pose exhibited by the detected face falls, the system comprising:

a general purpose computing device; and

a computer program comprising program modules executable by the computing device, wherein the computing device is directed by the program modules of the computer program to:

create database comprising a plurality of training feature characterizations, each of which characterizes the face of a person at a known face pose or a non-face;

train a plurality of detectors arranged in a pyramidal architecture to determine whether a portion of an input image depicts a person's face having a face pose falling within a face pose range associated with one of the detectors using the training feature characterizations; and wherein

said detectors using a greater number of feature characterizations are arranged at the bottom of the pyramid, and

said detectors arranged to detect finer ranges of face pose are arranged at the bottom of the pyramid; and wherein the program module to train a plurality of detectors comprises sub-modules to,

(a) input a set of training examples, a prescribed maximum number of weak classifiers, a cost function capable of measuring the overall cost, and an acceptable maximum cost;

(b) compute a set of weak classifiers, each classifier being associated to a particular feature of the training examples,

(c) determine which of the set of weak classifiers is the most significant classifier;

(d) add said most significant classifier to a current set of optimal weak classifiers;

(e) determine which of the current set of optimal weak classifiers is the least significant classifier;

(f) compute the overall cost for the current set of optimal weak classifiers using the cost function;

(g) conditionally remove the least significant classifier for the current set of optimal weak classifiers;

(h) compute the overall cost for the current set of optimal weak classifiers less the least significant classifier using the cost function;

(i) determine whether the removal of the least significant classifier results in a lower overall cost;

(j) whenever it is determined that the removal of the least significant classifier results in a lower overall cost, eliminate the least significant classifier;

(k) recompute each classifier in the current set of optimal weak classifiers associated with a feature added subsequent to the eliminated classifier while keeping the earlier optimal weak classifiers unchanged;

(l) repeat actions (f) through (k) until it is determined the removal of the least significant classifier does not result in a lower overall cost and then reinstate the last identified least significant classifier to the current set of optimal weak classifiers;

(m) determine if the number of weak classifiers in the current set of optimal weak classifiers equals the prescribed maximum number of weak classifiers or the last computed overall cost for the current set of optimal weak classifiers is less than the acceptable maximum cost; and

(n) whenever it is determined that the number of weak classifiers in the current set of optimal weak classifiers does not equal the prescribed maximum number of weak classifiers and the last computed overall cost for the current set of optimal weak classifiers exceeds the acceptable maximum cost, repeat actions (c) through (m) until it is determined that the number of weak classifiers in the current set of optimal weak classifiers does equal the prescribed maximum number of weak classifiers or the last computed overall cost for the current set of optimal weak classifiers becomes less than the maximum allowable cost, then output the sum of the individual weak classifiers as the trained strong classifier.

16. A computer-readable medium having computer-executable instructions for boosting the performance of a classifier in a statistical based machine learning system, said computer executable instructions comprising:

identifying a set of weak classifiers each of which is associated with a feature found in a plurality of training examples, said weak classifiers collectively best classifying the training examples;

linearly combining each of the weak classifiers in the identified set of weak classifiers to define a strong classifier,

wherein the action of identifying the set of weak classifiers comprises using a sequential forward search for optimal weak classifiers with backtracking to ensure the inclusion of a weak classifier in the set of weak classifiers in lower overall performance in the form of increased processing time.